POLYCYSTIC KIDNEY DISEASE *PKD2* GENE AND USES THEREOF

Statement of Government Interest

This invention was made with government support under NIH Grant Nos. DK48383 and DK02015. As such, the government has certain rights in this invention.

Background of the Invention

This invention is based upon the discovery by the inventors of the *PKD2* gene associated with Autosomal Dominant Polycystic Kidney Disease ("ADPKD"), the "*PKD2* gene" or "*PKD2*", and a novel protein encoded by this gene. The discovery of the *PKD2* gene and the protein encoded by the gene will have important implications in the diagnosis and treatment of ADPKD caused by defects in the *PKD2* gene.

ADPKD is a genetically heterogeneous disorder that affects approximately 500,000 Americans and five million individuals world wide, and accounts for 8 to 10% of all end stage renal disease (ESRD) worldwide (Gabow, P.A. N. Eng. J. Med. 329:332 (1993)). Its principal clinical manifestation is bilateral renal cysts that result in chronic renal failure in about 45% of affected individuals by age 60 (Gabow, P.A., supra). Hypertension and liver cysts are common, and the involvement of other organ systems (Gabow, P.A., et al. Kidney Int. 38:1177 (1990); Chapman, A.B., et al. N. Eng. J. Med. 327:916 (1992); Hossack, K.F., et al. N. Eng. J. Med. 319:907 (1988); Torres, V.E., et al. Am. J. Kidney Dis. 22:513 (1993); Huston, J., et al. J. Am. Soc. Nephrol. 3:1871 (1993); Somlo, S., et al. J. Am. Soc. Nephrol. 4:1371 (1993)) lends support to the view that polycystic kidney disease is a systemic disorder (Gabow, P.A., supra).

To date, most forms of ADPKD have been associated with two genes, *PKD1* and *PKD2*. The full genomic structure and cDNA sequence for the *PKD1* gene has been identified (The International Polycystic Kidney Disease Consortium, Cell

81:289 (1995); The American PKD1 Consortium, Hum. Mol. Genet. 4:575 (1995)). Mutations in the *PKD1* gene are suspected of causing 80-90% of all cases of ADPKD. The *PKD2* gene has been localized on chromosome 4q21-23 and accounts for approximately 15% of affected families (Kimberling, W.J., et al. Genomics 18:467 (1993); Peters, D.J.M. and L.A. Sandkuijl Contrib. Nephrol. 97:128 (1992)). Prior to the present invention, however, the *PKD2* gene had not been identified.

Summary of the Invention

The present invention provides a purified and isolated wild type *PKD2* gene, as well as mutated forms of this gene. The present invention also provides one or more single-stranded nucleic acid probes which specifically hybridize to the wild type *PKD2* gene or the mutated *PKD2* gene, and mixtures thereof, which may be formulated in kits, and used in the diagnosis of ADPKD associated with the mutated *PKD2* gene.

The present invention also provides a vector comprising nucleic acid encoding an active PKD2 protein, a cell stably transformed with this vector, as well as a method for producing recombinant, active PKD2 protein. A purified, active PKD2 protein is also provided by the present invention. In addition, the present invention provides an antibody immunoreactive with a wild type PKD2 protein, as well as an antibody immunoreactive with a mutant PKD2 protein, which may be formulated in kits, and used in the diagnosis of ADPKD associated with the mutated *PKD2* gene.

The present invention further provides a method for diagnosing ADPKD caused by a mutated *PKD2* gene in an adult subject suspected of having the disease comprising detecting the presence of a mutated *PKD2* gene in nucleic acid of the subject. The present invention still further provides a method for treating ADPKD caused by a mutated *PKD2* gene in a subject in need of such treatment comprising the delivery and

expression of a functional *PKD2* gene into a sufficient number of cells of the subject to treat the disease. A stem cell which expresses the *PKD2* gene introduced therein through viral transduction, homologous recombination or transfection is also provided by the invention.

In addition, the present invention provides a recombinant viral vector for treating a defect in the *PKD2* gene in a target cell comprising (a) the nucleic acid of or corresponding to at least a portion of the genome of a virus, which portion is capable of directing the infection of the target cell, and (b) a *PKD2* gene operably linked to the viral nucleic acid and capable of being expressed as a functional gene product in the target cell.

Finally, the present invention provides a vector and an embryonic stem cell each of which comprises a mutated *PKD2* gene, a non-human, transgenic animal whose germ and somatic cells contain a mutated *PKD2* gene sequence introduced into said animal, or an ancestor thereof, at an embryonic stage, as well as a method for producing the non-human, transgenic animal.

Additional objects of the invention will be apparent from the description which follows.

Brief Description of the Figures

Figure 1A represents the subset of STSs from the high density map of the *PKD2* region showing polymorphic loci flanking the interval. JSTG3 and AICA1 are two of nine microsatellite markers in this region developed previously. *SPP1* (osteopontin, STS4-1078) and D4S1171 were used to screen the P1 library as described in Materials and Methods. Other sources of STSs include published linkage maps and genome center databases. cen, centromere; tel telomere. Distances are in Morgans along chromosome 4.

Figure 1B shows representative mega-YACs (Cohen, D., et al. Nature 366:698 (1993)), and their STS content, forming a contig around the *PKD2* region.

Figure 1C shows the minimum tiling path of the cosmid and P1 contig in the *PKD2* region. Clone names beginning with "c" and "p" refer to cosmid and P1 clones, respectively; addresses are from the original arrayed libraries. The clones containing JSTG3 and AICA1 are shown; a single gap of <40 kb is indicated by the arrow.

Figure 1D shows the detail of the portion of the contig containing the *PKD2* candidate gene, cTM-4.

Figure 1E shows overlapping map of nine cDNA clones for cTM-4 and a composite schematic at the bottom. Clones K1-1 and K1-5 are from the adult kidney library; clones yj63h09 and yc93g07 were identified by GenBank searching and are from the normalized infant brain library (Soares, M.B., et al. Proc. Natl. Acad. Sci. USA 91: 9228 (1994)); all other clones are from the fetal brain library. Shaded areas represent chimeric portions of clones.

Figure 2 represents expression of the *PKD2* candidate gene. Insert from cTM-4B3-3 (Figure 1E) was used as a hybridization probe on mRNA blots containing human tissues (Clonetech, Palo Alto, CA). Hybridization was performed without pre-competition and a final wash stringency of 0.5X SSC, 0.1% SDS at 65°C. Tissues in numbered lanes are: (1) heart, (2) brain, (3) placenta, (4) lung, (5) liver, (6) skeletal muscle, (7) kidney, (8) pancreas, (9) spleen, (10) thymus, (11) prostate, (12) testis, (13) ovary, (14) small intestine, (15) colon, (16) leukocytes, (17) fetal brain, (18) fetal lung, (19) fetal liver, (20) fetal kidney. At bottom, β-actin hybridization to the same blots is used to compare relative mRNA loading within each blot.

Figure 3 depicts the mutations in *PKD2* from an analysis of genomic PCR products in three *PKD2* families. Left panel shows the results of direct sequencing of genomic

PCR products from affected individuals. The arrows denote double peaks, confirmed by sequencing in both directions, indicative of heterozygosity at that nucleotide. Each of the mutant alleles results in a premature stop codon. The right panel demonstrates segregation of the mutated allele with the disease phenotype. In families 97 and 1605, the affected alleles are not digested by Bsr I and Taq I, respectively, since the restriction sites are lost by mutation. Family 1601 shows segregation of the SSCA variant, indicated by the arrow, with the disease phenotype. For each family, only portions of more extensive pedigrees are shown. Filled symbols, affected individuals. Open symbols, unaffected individuals. M, 100 bp ladder.

Figure 4A depicts the deduced amino acid sequence of PKD2 (cTM-4) (GenBank accession: gb|U50928) in alignment with PKD1 (gb|U24497), the *C. elegans* homolog of PKD1 (ZK945.9; swiss|Q09624) and VACCα1E-1 (pir|B54972) using BESTFIT (Program Manual for the Wisconsin Package, Version 8, September 1994): identity to cTM-4, |; similarity to cTM-4, :. Numbers in parentheses refer to amino acids in respective sequences. Putative transmembrane domains, tm1 to tm6. Predicted *N*-glycosylation sites, *. Potential phosphorylation sites with strong consensus sequences: protein kinase C, +; cGMP dependent kinase, open square (Ser 826 is also consistent with a protein kinase A site); casein kinase, open circle. The sites of the nonsense mutations (Fig. 3) are indicated by arrows labeled with the respective family numbers. The EF-hand domain is indicated by the dashed line.

Figure 4B shows alignment of the EF-hand domain with the EF-hand test sequence. The residues E, G, I, and E, the latter being a Ca^{2+} coordination vertex, are the expected residues at the indicated positions in the EF-hand. Positions indicated as "n" are expected to have hydrophobic amino acids (L, I, V, F, M); those denoted with * should be

oxygen-containing amino acids (D, N, E, Q, S, T) comprising the remainder of coordination vertices for Ca^{2+} binding; the -Y vertex can be any amino acid. The Leu (L) in PKD2 in place of the Ile (I) is likely a permissible substitution; PKD2 has Gln (Q) in place of the consensus Gly (G) as is the case with EF-hand domains in the α1 Na^+ channels.

Figure 5 represents the nucleotide sequence of the *PKD2* gene and the deduced amino acid sequence for PKD2.

Detailed Description of the Invention

The present invention provides a purified and isolated wild type *PKD2* nucleic acid, as well as mutated forms of this nucleic acid. As used herein, the "wild type *PKD2* nucleic acid" is the normal form of the gene which expresses an enzymatically active gene product, and includes degenerate forms. The "mutated *PKD2* nucleic acid" is the mutated form of the normal *PKD2* gene, which contains one or more deletion, insertion, point or rearrangement mutations, or a combination thereof, that may render the gene product expressed by the mutated *PKD2* gene nonfunctional or nonexistent. As used herein, "nucleic acid" may be genomic DNA, cDNA or RNA.

The present invention also provides single-stranded nucleic acid probes and mixtures thereof for use in diagnosing ADPKD caused by a mutated *PKD2* gene. The nucleic acid probes may be DNA, cDNA, or RNA, and may be prepared from the mutated and/or wild type *PKD2* gene. The probes may be the full length sequence of *PKD2* gene, or fragments thereof. Typical probes are 12 to 40 nucleotides in length. Generally, the probes are complementary to the *PKD2* gene coding sequences, although probes to introns are also contemplated. The probes may be synthesized using an oligonucleotide synthesizer such as Applied Biosystems Model 392 DNA/RNA synthesizer, and may be labeled with a detectable marker such as a fluorescence, enzyme or radiolabeled markers

including ^{32}P and biotin, and the like. Combinations of two or more labelled probes corresponding to different regions of the *PKD2* gene also may be included in kits to allow for the detection and/or analysis of the *PKD2* gene by hybridization.

The present invention also provides a vector comprising nucleic acid encoding an active PKD2 protein, as well as a cell stably transformed with the vector. The vector may be any plasmid, viral-derived nucleic acid, lytic bacteriophage derived from phage lambda, cosmid, filamentous single-stranded bacteriophage such as M13, and the like, for cloning nucleic acid or introducing the nucleic acid into a cell for expression. The cell may be eukaryotic or prokaryotic. Suitable host cells include but are not limited to bacterial cells such as E. coli, Bacillus subtilis, Agrobacterium tumefaciens, Bacillus subtilis, Agrobacterium tumefaciens, Bacillus megaterium, eukaryotic cells such as Pichia pastoris, Chlamydomonas reinhardtii, Cryptococcus neoformans, Neurospora crassa, Podospora anserina, Saccharomyces cerevisiae, Saccharomyces pombe, Uncinula necator, cultured insect cells, cultured chicken fibroblasts, cultured hamster cells, cultured human cells such as HT1080, MCF7, 143B and cultured mouse cells such as EL4 and NIH3T3 cells. Such expression systems may be used to produce a recombinant, active PKD2 protein by culturing a cell transformed with a vector comprising a nucleic acid encoding an active PKD2 protein, and recovering PKD2 protein from the culture.

The present invention also provides a purified active PKD2 protein. The protein may be the wild type protein or an analogue thereof. As used herein, "analogue" means functional variants of the wild type protein, and includes PKD2 proteins isolated from mammalian sources other than human, as well as functional variants thereof. The protein also may be isolated from native cells or recombinantly produced.

The present invention also provides antibodies immunoreactive with the protein expressed by the wild type *PKD2* gene (and analogues thereof), as well as antibodies immunoreactive with the protein expressed by the mutated *PKD2* gene. The antibodies may be polyclonal or monoclonal and are produced by standard techniques. The antibodies may be labeled with standard detectable markers (e.g. chemiluminescent detection systems and radioactive labels such as ^{125}I) for detecting the wild type and mutated *PKD2* genes. The antibodies also may be presented in kits with detectable labels and other reagents and buffers for such detection.

The present invention also provides a method for diagnosing ADPKD in a subject comprising detecting the presence of a mutated *PKD2* gene in nucleic acid of the subject. The method may be used to determine whether persons in the population at large have ADPKD, for identifying persons at risk in developing the disease, i.e. relatives of persons with ADPKD, as well as for confirming diagnosis of ADPKD. The method also is useful for diagnosing ADPKD before clinical manifestations of the disease, i.e. the formation of cysts. Accordingly, as used herein, "subject" may be an embryo, fetus, newborn, infant or adult.

The presence of the mutated *PKD2* gene may be detected by procedures known in the art including but not limited to standard sequencing techniques (e.g. dideoxy chain termination), restriction enzyme digestion analysis, hybridization with one or more probes hybridizable to the mutated and/or wild type *PKD2* gene using standard procedures such as Southern blot analysis, polymerase chain reaction using sense and antisense primers prepared from the mutated and/or wild type *PKD2* genes, and combinations thereof.

The presence of the mutated *PKD2* gene also may be detected by detecting expression of the gene product of the gene. Such expression products include both mRNA as well as

the protein product itself. mRNA expression may be detected by standard sequencing techniques, hybridization with one or more probes hybridizable to the mutated and/or wild type *PKD2* mRNA using standard procedures such as Northern blot analysis, dot and slot hybridization, S1 nuclease assay, or ribonuclease protection assays, polymerase chain reaction using sense and antisense primers prepared from the mutated and/or wild type *PKD2* genes, and combinations thereof. The protein may be detected using antibodies to the protein expressed by the mutated *PKD2* gene and/or the wild type *PKD2* gene by procedures known in the art including but not limited to immunoblotting, immunoprecipitation, solid phase radioimmunoassay (e.g. competition RIAs, immobilized antigen or antibody RIAs, or double antibody RIAs), enzyme-linked immunoabsorbent assay, and the like.

The present invention also provides a method for treating ADPKD caused by a mutated *PKD2* gene in a subject in need of such treatment comprising the delivery and expression of a functional *PKD2* gene into a sufficient number of cells of the subject, preferably bone marrow stem cells, to treat ADPKD in the subject. As used herein, "functional *PKD2* gene" is a gene which when incorporated into a cell's nucleic acid expresses a functional gene product, and includes the wild type *PKD2* gene as well as variations thereof. The delivery and expression of the functional *PKD2* gene may be accomplished by introducing the functional *PKD2* gene into the cells or by correcting the mutation(s) in the subject's *PKD2* gene.

The functional *PKD2* gene may be delivered into the subject's cells by a number of procedures known to one skilled in the art, e.g. electroporation, DEAE dextran, cationic liposome fusion (using both monocationic and polycationic lipids), protoplast fusion, DNA coated microprojectile bombardment, injection with recombinant replication-defective retroviruses, homologous recombination,

and the like. Accordingly, a stem cell which expresses the *PKD2* gene introduced therein through viral transduction, homologous recombination, or transfection is also provided by the present invention.

The present invention also provides a recombinant viral vector for treating a defect in the *PKD2* gene in a target cell comprising (a) the nucleic acid of or corresponding to at least a portion of the genome of a virus, which portion is capable of directing the infection of the target cell, and (b) a functional *PKD2* gene operably linked to the viral nucleic acid and capable of being expressed as a functional gene product in the target cell. The recombinant viral vectors of the present invention may be derived from a variety of viral nucleic acids known to one skilled in the art, e.g. the genomes of HSV, adenovirus, adeno-associated virus, Semiliki Forest virus, vaccinia virus, and other retroviruses or DNA viruses.

The present invention also provides a vector for use in preparing a non-human, transgenic animal comprising a mutated *PKD2* gene which is capable of introducing the mutated *PKD2* gene in at least some embryonic cells to which the vector is introduced, an embryonic stem cell comprising a mutated *PKD2* gene which has been integrated into the cell following transduction with the vector above, as well as a non-human transgenic animal of ADPKD which would be useful for studying ADPKD. The mutated *PKD2* gene may be integrated into the germ line of a non-human animal such as a mouse, rat, goat, sheep or other non-human species in order to obtain a transgenic animal model by methods known in the art (see Alberts, B., et al. Molecular Biology of the Cell, 2d. Garland Publ. Inc., New York and London, pp. 267-269 (1989)). For example, nucleic acid encoding the mutated PKD2 protein can be inserted into the genome of a replication-defective virus such as HSV or a retrovirus or transposen and the resultant construct injected into embyronic stem cells.

Alternatively, the transgenic animal may be made by injecting nucleic acid into the male pronucleus of a fertilized egg of a nonhuman animal, transplanting the "transgenic embryo" into a pseudopregnant female and then analyzing offspring for the presence of the injected nucleic acid in their genome.

The present invention is described in the following Experimental Details Section, which is set forth to aid in an understanding of the invention, and should not be construed to limit in any way the invention as defined in the claims which follow thereafter.

Experimental Details Section

A. Materials and Methods

Cosmid and P1 Contig Construction. Cosmid and P1 contig construction was guided by the existing YAC contig (Mochizuki, T., et al., unpublished observations; Veldhuisen, B., et al., unpublished observations). Cosmid clones were obtained by hybridization screening of the human chromosome 4-specific cosmid library (Riess, O., et al. Cytogenet. Cell Genet. 65:238 (1994); Ioannou, P.A., et al. Nature Genetics 6:84 (1994)). The probes used for hybridization were: 1) pooled *Alu* products from mega-YAC 967d1, 2) end sequences from cosmid or P1 clones mapping into the region, and 3) internal restriction fragments from the YAC, P1 and cosmid clones. All hybridization probes were α^{32}P-dCTP labeled by standard techniques. Human repetitive sequences were pre-competed with 30-80 μg of C_0t-1 DNA using the manufacturer's portocol (Gibco/BRL, Gaithsburg, MD). Hybridization was carried out in Church-Gilbert buffer. The P1 library (Riess, O., et al., supra; Ioannou, P.A., et al., supra) was screened by PCR from colony pools of each 384-well plate using STS4-1078 (*SPP1*) and D4S1171. Cosmid and P1 clones mapping into the PKD2 interval were screened for STS content to anchor positive clones onto the YAC contig. Overlap relationships

among the clones were established by Eco RI fingerprint analysis and by hybridization.

Screening of cDNA Clones. $6X10^5$ plaques of oligo-dT and random primed human fetal brain (Stratagene #936206) and adult kidney (Clonetech #HL3001a) cDNA libraries were plated at a density of $3X10^4$ per 150 mm plate and replica lifted onto nylon filter circles. Cosmid and P1 inserts used in library screening were released from vector with Not I and purified from agarose gels. The cumulative length of inserts used as probe in a library screening was <80 kb to maintain adequate signal-to-noise. Insert DNA was labeled and pre-competed with 2 μg sCOS-1 vector in addition to C_0t-1 DNA. Positively hybridizing plaques were purified by standard techniques and insert DNA was excised (λZAPII) or subcloned (λgt10).

Identification of Mutations. The mutation in family 1605 was detected initially in RT-PCR template using the cDNA-based primers F11 (5'-GGGCTACCATAAAGCTTTG-3') and R11 (5'-GTTCATGTTCGATCAGTTCT-3') (205 bp product) and confirmed in genomic DNA using F11 with intronic primer IR11 (5'-GGGCTAGAAATACTCTTATCACC-3') (201 bp product). The mutations in families 97 and 1601 were initially detected in genomic DNA using intronic primers IF1C (5'-GCCTCAAGTGTTCCACTGAT-3') and IR1 (5'-AGGTTTTTCTGGGTAACCCTAG-3') (362 bp product). Amplifications were performed in standard conditions with hot start. Products were labeled by α^{32}P-dCTP incorporation, diluted and denatured in formamide buffer prior to electrophoresis. SSCA was performed according to published protocols (Orita, M., et al. Genomics 5:874 (1989)). Sequencing of purified PCR products was performed with either an ABI 373a or 377 automated sequencing apparatus using cycle sequencing with dye terminator chemistries according to the manufacturer's protocol. The PCR primers were used as

sequencing primers and all products were sequenced in both directions. The mutation in family 97 results in the loss of a Bsr I site. Genomic DNA amplified with IF1C and IR1 and digested with Bsr I yields products of 261 and 101 bp in the normal allele. The mutation in family 1605 results in the loss of a Taq I site. Genomic DNA amplified with F11 and IR11 and digested with Taq I yields products of 105 and 96 bp in the normal allele. The SSCA conditions used to demonstrate the mutation in the IF1C-IR1 genomic PCR product in family 1601 were 6% acrylamide (29:1), 1X TBE, on a 20 cm gel run at 14°C and 100 V for 6 hours.

B. Discussion

The *PKD2* genetic interval is flanked by the polymorphic markers D4S231 and D4S414/423 (Kimberling, W.J., et al. Genomics 18:467 (1993); Peters, D.J.M., et al. Nature Genetics 5:359 (1993)). A yeast artificial chromosome (YAC) contig and high density sequence tag site (STS) map of this region was constructed as described above (Figure 1). Genetic studies in affected families using physically ordered polymorphic markers led to several progressive refinements of the *PKD2* interval (Mochizuki, T., et al. J. Am. Soc. Nephrol. 5:631a (1994); San Millian, J.L., et al. Am. J. Hum. Genet. 56:248 (1995); Peters, D.J.M., et al. Am. J. Hum. Genet. 57:200a (1995); Constantinou-Deltas, C.D., et al. Hum. Genet. 95:416 (1995)). The closest unambiguous flanking genetic markers are AFMa059xc9 proximally and AICA1 distally (Figures 1A, B). A cosmid- and P1-based (Riess, O., et al. supra; Ioannou, P.A., et al., supra) contig extending over ~680 kb from AICA1 to the region centromeric to the polymorphic marker JSTG3 was constructed as described above (Figure 1C). This contig contains a single gap of less than 40 kb. cDNAs corresponding to genes in this region were isolated using inserts from the genomic clones to screen either a human fetal brain or adult kidney cDNA library as described above.

The mapping of the cDNA clones identified was confirmed and the clones were sequenced. These sequences were analyzed to identify open reading frames (ORF) and database searches using the BLAST algorithms (Altschul, S.F., et al. J. Mol. Biol. 215:403 (1990)) were performed.

One group of clones, collectively termed cTM-4, were initially isolated using insert DNA from cosmid c44a9 from the chromosome 4-specific cosmid library as described above (Figure 1D). None of the cTM-4 clones have nucleotide level homology to any known genes, although 2 randomly-sequenced cDNA clones were identified (Figure 1E). Northern blot hybridization with the cTM-4B3-3 insert (Figure 1E) revealed a ~5.4 kb transcript expressed in most fetal and adult tissues (Figure 2). cTM-4 is strongly expressed in ovary, fetal and adult kidney, testis, small and large intestine, and fetal lung. Peripheral blood leukocytes was the only tissue tested in which expression was not detected.

Initial database searching using the 6 translated reading frames obtained from the sequence of clone cTM-4B3-3 revealed amino acid level homology with polycystin, the *PKD1* gene product (The European Polycystic Kidney Disease Consortium, Cell, 77:881 (1994); The International Polycystic Kidney Disease Consortium, Cell 81:289 (1995); The American PKD1 Consortium, Hum. Mol. Genet. 4:575 (1995); Hughes, J., et al. Nature Genetics 10:151 (1995)). Based on its map location, pattern of expression and the observed homology, the cTM-4 gene was further investigated as a candidate for *PKD2*. Nine overlapping cDNA clones were completely sequenced in both directions (Figure 1E). The 5' end of the cTM-4 gene contains a pair of genomic Not I sites and the 3' end extends in the telomeric direction beyond the end of the P1 clone p157n2, into cosmid c44a9 (Figures 1C, 1D). The gene extends over 68 kb of the genome.

The consensus 5057 bp sequence (GenBank accession: gb|U50928) is represented schematically in Figure 1E. A

translation start site with a good Kozak consensus sequence (5'-ACCGCGATGG-3') (Kozak, M. Nucleic Acids Res. 15:8125 (1987)) was identified 67 bp from the 5' end of the K1-1 clone and 61 bases after an in-frame stop codon. It is followed by a 2904 bp ORF followed, in turn, by several in-frame stop codons. The 3' untranslated region is 2086 bp long and contains a consensus polyadenylation signal.

The DNA sequence and expression profiles of cTM-4 was next analyzed in unrelated affected individuals from families with PKD2 (Kimberling, W.J., et al. N. Eng. J. Med. 319:913 (1988); Kimberling, W.J., et al. Genomics 18:467 (1993); Peters, D.J.M., et al. Nature Genetics 5:359 (1993); Constantinou-Deltas, C.D., et al. Hum. Genet. 95:416 (1995)). Reverse transcribed RNA and genomic DNA templates were used to generate PCR products for single strand conformational analysis (SSCA) as described above. Genomic PCR products of SSCA variants identified in three families were subjected to direct sequencing. Each affected individual was found to be heterozygous for a single base change that resulted in a nonsense mutation (Figure 3). The mutation in family 97 is a G to A transition in the codon for Trp 380 (Figures 3, 4). The mutations in the Cypriot families 1605 and 1601 are C to T transitions in codons Arg 742 and Gln 405, respectively (Figures 3, 4). Using either the resultant loss of a restriction site in families 97 and 1605, or the SSCA pattern in family 1601, segregation of the mutation with the disease phenotype in each family was demonstrated (Figure 3). Analysis of between 90 and 100 normal chromosomes failed to show the predicted affected allele in any case, making it less likely that these sequence differences represent anonymous polymorphisms. These limited findings do not provide evidence for clustering of mutations in *PKD2*.

The identification of mutations that disrupt the predicted translation product of cTM-4 and the segregation of these mutations with the ADPKD phenotype in three well

characterized PKD2 pedigrees, provide strong evidence that cTM-4 is the *PKD2* gene. The putative translation product of the cTM-4 ORF is a 968 amino acid sequence with a calculated molecular mass of 110 kD. Modeling with several hydrophobicity algorithms (Rost, B., et al. Protein Sci. 4:521 (1995); Klein, P., et al. Biochim. Biophys. Acta 815:468 (1985); Kyte, J. and R.F. Doolittle J. Mol. Biol. 157:105 (1982); Engelman, D.M., et al. Annu. Rev. Biophys. Chem. 15:321 (1986)) suggest that cTM-4 is an integral membrane protein with six (range, 5 to 8) membrane spanning domains and intracellular NH_2- and COOH-termini (Sipos, L. and G. von Heijne Eur. J. Biochem. 213:1333 (1993); Nakashima, H. and K. Nishikawa FEBS Lett. 303:141 (1992); Hartmann, E., et al. Proc. Natl. Acad. Sci. USA 86:5786 (1989)). Of the six highest scoring domains, the fourth transmembrane domain (tm4, Figure 4), produced the lowest scores, but was consistently predicted to be a membrane span by several analyses (Rost, B., et al., supra; Klein, P., et al., supra; Kyte, J. and R.F. Doolittle, supra; Engelman, D.M., et al., supra). The "inside positive" rule (Sipos, L. and G. von Heijne, supra; Nakashima, H. and K. Nishikawa, supra; Hartmann, E., et al., supra) strongly supports the predicted topology. The majority of the *N*-glycosylation sites, occurring in the segment between tm1 and tm2 (Figure 4), are predicted to be extracellular. In addition, potential phosphorylation sites were identified primarily in the COOH-terminal region, as was a putative EF-hand domain (Kretsinger, R.H. Cold Spring Harb. Symp. on Quant. Biol. 52:499 (1987); Babitch, J. Nature 346:321 (1990)), and this region is predicted to be intracellular (Figure 4). If a stable protein product is produced, the mutations in families 97 and 1601 are expected to result in a product with an intact intracellular NH_2-terminal domain, first transmembrane domain, and part of the first extracellular loop. The mutation in family 1605 is predicted to result in a product

lacking the portion of the intracellular COOH-terminus that contains several phosphorylation sites and the EF-hand domain (Figure 4).

There is ~25% identity and ~50% similarity between the putative translation product of *PKD2* and ~450 amino acids of polycystin and its *C. elegans* homolog, ZK945.9 (Figure 4). There is a comparable degree of similarity with ~270 residues of the voltage activated Ca^{2+} channel α_{1E} (VACCα1E-1; Figure 4). The similarity between PKD2 and polycystin (and ZK945.9) extends over the region tm1 to tm6 in PKD2 but does not include the NH_2- and COOH-terminal domains. The corresponding region of polycystin has been predicted to contain four transmembrane segments (Hughes, J., et al. Nature Genetics 10:151 (1995)), three of these corresponding to tm1, tm2 and tm5 in the *PKD2* gene product and the fourth localizing between tm5 and tm6 of PKD2. The regions corresponding to tm3 and tm4 of PKD2 were not predicted to be membrane spans in that report (Hughes, et al., supra).

The similarity to VACCα1E-1 (Williams, M.E., et al. J. Biol. Chem. 269:22347 (1994); Williams, M.E., et al. Science 257:389 (1992); Soong, T.W., et al. Science 260:1133 (1993); Horne, W.A., et al. Proc. Natl. Acad. Sci. USA 90:3787 (1993)) is presented as the strongest example of a general homology of PKD2 to the family of voltage activated Ca^{2+} and Na^{+} α1 channel proteins. These channel proteins contain four homologous domains (I-IV), each with six transmembrane spans (S1-S6), which are predicted to form the pore structure (Williams, M.E., et al. (1994), supra; Williams, M.E., et al. (1992), supra; Soong, T.W., et al., supra; Horne, W.A., et al., supra; Hille, B., Ionic channels of excitable membranes (Sinauer Associates, Sunderland, MA, ed. 2, 1992), pp.250-7)). The membrane spans tm2 through tm6 as well as the intervening intracellular loops of PKD2 have similarity with corresponding segments in the α1 channels (Figure 4). The similarity in the COOH-terminal region

includes the putative EF-hand domain (Figure 4) (Kretsinger, R.H., et al., supra; Babitch, J., supra). This domain in PKD2 scores highly on the EF-hand test (Figure 4B) with identity at all the critical coordination vertices (Kretsinger, R.H., et al., supra; Babitch, J., supra). EF-hand domains are specialized helix-loop-helix motifs that have Ca^{2+} binding activity in ~70% of proteins in which they occur (Nakayama, S. and R.H. Kretsinger Annu. Rev. Biophys. Biomol. Struct. 23:473 (1994)). Unpaired EF-hand sequences have recently been implicated in Ca^{2+}-sensitive inactivation of some forms of L-type VACCα1 (de Leon, M., et al. Science 270:1502 (1995)). EF-hand domains that do not coordinate Ca^{2+} remain important to protein function (Kretsinger, R.H., et al., supra; Babitch, J., supra; Gulati, A., et al. J. Biol. Chem. 267:25073 (1992)).

Despite the observed homology to PKD1, the predicted structure of the PKD2 protein does not directly suggest a role in cell-cell or matrix-cell signaling similar to that proposed for polycystin (The European Polycystic Kidney Disease Consortium, supra; The International Polycystic Kidney Disease Consortium, supra; The American PKD1 Consortium, supra; Hughes, J., supra). PKD2 does not have the large NH_2-terminal extracellular domain and the associated motifs found in polycystin (The European Polycystic Kidney Disease Consortium, supra; The International Polycystic Kidney Disease Consortium, supra; The American PKD1 Consortium, supra; Hughes, J., supra). It is possible that PKD2 functions in a parallel pathway with PKD1. However, given that the clinical diseases produced by mutations in *PKD1* and *PKD2* exhibit an identical spectrum of organ involvement, differing only in relative rates of progression of cystic changes, hypertension and the development of ESRD, the most likely scenario is that PKD2 associates with itself, with polycystin, and/or with other

proteins and ligands as part of a common signal transduction pathway.

PKD2 bears some similarity to the α1 Ca^{2+} (and Na^{+}) channels but has only six membrane spans. If it formed homo- or hetero-multimeric complexes (for example, with itself, with the homologous portion of PKD1 or with another protein), it could function as an ion channel or pore in a manner similar to the K^{+} channels (Hille, B., supra). The observed homologies, the presence of a pair of conserved basic residues (Lys 573, Lys 576) in the fourth transmembrane domain, and the predicted even number of membrane spans, are consistent with such a role (Hille, B., supra). In such a model, PKD1 could act as the regulator of the PKD2 channel activity, perhaps with Ca^{2+} as a second messenger in a signal transduction pathway. The discovery of PKD2 raises the possibility that the ADPKD phenotype may in part be the result of a defect in an unknown transport function.

All publications mentioned hereinabove are hereby incorporated by reference in their entirety.

While the foregoing invention has been described in detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of the disclosure that various changes in form and detail can be made without departing from the true scope of the invention in the appended claims.